Ejercicio adicional de regresión para resolver en el Laboratorio

A mediados del siglo XIX, un físico escocés, James D. Forbes, quiso determinar la altitud sobre el nivel del mar a través de la medición del punto de ebullición del agua. Él sabía que la altitud podía determinarse a través de la presión atmosférica, ya que mayor altitud, menor presión. Sin embargo, debido a la dificultad de transportar los frágiles barómetros de la época, él realizó experimentos en los Alpes y en Escocia con el objetivo de ver cómo se relacionan la presión (Y, medida en pulgadas de mercurio) y el punto de ebullición (X, medido en grados Fahrenheit). Forbes se hizo las siguientes preguntas: ¿cómo se relacionan la presión y el punto de ebullición?, ¿puede predecirse la presión a partir de la temperatura del punto de ebullición?.
Los datos, publicados en un trabajo de Forbes en 1857, dados en el archivo *forbes.xls*, son los siguientes:

punto de ebullición (en ºF)	**presión** (en Hg)
194,5	20,79
194,3	20,79
197,9	22,40
198,4	22,67
199,4	23,15
199,9	23,35
200,9	23,89
201,1	23,99
201,4	24,02
201,3	24,01
203,6	25,14
204,6	26,57
209,5	28,49
208,6	27,76
210,7	29,04
211,9	29,88
212,2	30,06

a) Represente estos datos en un diagrama de dispersión. ¿Es razonable suponer que existe una relación lineal entre *X* e *Y* que permita predecir *X* en función de *Y*?
b) Ajuste de la recta de cuadrados mínimos considerando las variables X e Y.
   i) Estudie la normalidad de los residuos.
   ii) En base al análisis realizado en i), ¿son válidas las hipótesis del modelo al ajustar X vs. Y?
c) En su trabajo, Forbes sugiere que si se grafican la presión versus el logaritmo (en base 10) del punto de ebullición multiplicado por 100 se obtiene una recta. Realice el diagrama de dispersión de X vs. Z=100*log(Y) ¿está de acuerdo con Forbes?.
d) Ajuste de la recta de cuadrados mínimos considerando las variables X y Z.
   i) Estudie la normalidad de los residuos a partir de los plots de
      (1) probabilidad de los residuos estandarizados.
      (2) residuos estandarizados vs. valores ajustados.
   ii) A partir de i), se puede ver que existe un dato atípico; determine cuál es.
e) Ajuste de la recta de cuadrados mínimos considerando las variables X y Z, pero sin considerar el dato atípico
   i) Estudie la normalidad de los residuos a partir de los plots de:
      (1) probabilidad de los residuos estandarizados.
      (2) residuos estandarizados vs. valores ajustados.
   ii) ¿Se verifican las hipótesis del modelo?
f) A partir del ajuste realizado en e):

i) testee la hipótesis $H_0 : z_i = \alpha + e_i$, i = 1,…,n vs $H_1 : z_i = \alpha + \beta x_i + e_i$, i = 1,…,n.

ii) determine la pendiente y la ordenada al origen de la recta de cuadrados mínimos.

iii) testee la hipótesis $H_0 : \alpha = 0$, β arbitrario vs $H_0 : \alpha \neq 0$, β arbitrario.

iv) testee la hipótesis $H_0 : \beta = 0$, α arbitrario vs $H_0 : \beta \neq 0$, α arbitrario.

v) encuentre un intervalo de confianza del 95% para la pendiente. Idem para la ordenada al origen.

vi) grafique los intervalos de confianza predicción y los intervalos de confianza para el valor esperado de Z.

vii) encuentre un intervalo de confianza del 95% para el valor esperado de Z cuando el punto de ebullición es 200º F

viii) Este año, un científico fue a los Alpes y repitió el experimento realizado por Forbes.

(1) Si una de sus mediciones de la presión resultó ser 25,5 Hg, encuentre un intervalo de confianza del 95% para el valor verdadero del punto de ebullición.

(2) Si una de sus mediciones del punto de ebullición fue 215º F, encuentre un intervalo de confianza del 95% para el predictor de la presión.